EARTH SURFACE PROCESSES AND LANDFORMS, VOL 22, 1061–1077 (1997)

BIAS AND PRECISION OF PERCENTILES OF BULK GRAIN SIZE DISTRIBUTIONS

R. I. FERGUSON[1] AND C. PAOLA[2]

[1]*Department of Geography, University of Sheffield, Sheffield S10 2TN, UK*
[2]*Department of Geology and Geophysics, University of Minnesota, Minneapolis MN 55455-0219, USA*

Received 9 September 1996; Revised 27 January 1997; Accepted 17 March 1997

ABSTRACT

Percentiles such as D_{50} and D_{84}, calculated from weights retained on different sieves, are widely used to characterize grain size distributions (GSDs) of bulk samples of sedimentary deposits or sediment fluxes. The sampling variability of such percentiles is not well known, and few sampling guidelines exist for reliable characterization of GSDs. We report results from computer sampling experiments on the variability of sample percentiles in different-sized samples from populations with a log-normal GSD by weight and different sorting coefficients. Sample sizes are scaled by the volume of a median-sized grain so that results can be applied to any log-normal GSD. Sampling is random for the GSD by number that is equivalent to a specified GSD by weight. Results show important differences from standard sampling theory applicable to pebble-count GSDs. In small bulk samples all percentiles, including the median, are underestimated (more so for smaller samples, coarser percentiles and poorer sorting), and precision does not improve with the square root of sample size until fairly large sample sizes are exceeded. Non-dimensional equations fitted by eye to the results give good approximations to expected bias and precision in any percentile from 50 to 95 for any given sample size and population sorting coefficient. They are inverted to estimate the sample size required to avoid significant bias, or achieve specified precision, in any percentile of interest given estimates of the population D_{50} and sorting coefficient. Target sample sizes are sometimes considerably smaller, but in other circumstances larger, than suggested by previous guidelines relating to estimation of the entire grain size distribution. Bias is likely in small samples of river bedload and good precision requires very large samples of poorly sorted gravel deposits.

Earth surf. process. landforms, **22,** 1061–1077 (1997)
No. of figures: 7 No. of tables: 1 No. of refs: 21
KEY WORDS: sampling; gravel; percentiles; bias; precision

INTRODUCTION

Grain size analysis is a standard tool in the study of sediments. Grain size distributions (GSDs) based on samples of sediment are widely used to describe parent populations, identify differences in space and time, and make inferences about process or origin. Often the GSD is represented by parameters such as the mean or standard deviation which are estimated from sample statistics. If the size of every grain in the sample is known, sampling variability in statistics such as the mean can be assessed using standard statistical theory. But usually what is known is either the number of particles, or the mass, in each of several sieve fractions, according to whether the GSD is by number (a pebble count) or by mass (a bulk sample). The GSD is then represented using interpolated percentiles of the sample, for example the median (D_{50}) instead of the mean, and D_{84} and D_{90} (sizes than which 84 per cent or 90 per cent is finer) to characterize the coarse tail. Very little is known about the sampling variability of GSD percentiles, or conversely the sample size required to give estimates of a specified precision. It is commonly thought that sample $D_{\max}$ (which can be seen as the 100th percentile) is a biased estimator of the true maximum grain size in the population, because a small sample is most unlikely to contain this unique grain and will therefore underestimate its size (e.g. Wilcock, 1992). However, we are unaware of published discussion of possible bias in other GSD percentiles, and suspect that most geomorphologists and sedimentologists assume they are unbiased and have precision that improves progressively with sample size. This appears to be so for pebble counts where the number n of grains in the sample is known. A binomial model can then be used to calculate the standard error of percentile $100p$ as:

$$s_p = \frac{\sigma}{y_p}\sqrt{\frac{p(1-p)}{n}} \tag{1}$$

* Correspondence to: R. I. Ferguson; (e-mail r.ferguson@sheffield.ac.uk)

CCC 0197-9337/97/111061–17 $17.50

where y_p is the ordinate of the population probability density function (PDF) of grain size at the chosen percentile, and σ the standard deviation of this PDF (e.g. Yule and Kendall, 1953). Rice and Church (1996) empirically confirmed this $1/\sqrt{n}$ improvement in precision, without assumptions about the PDF. But in bulk sampling the number of grains is unknown and is itself a stochastic variable. De Vries (1970) used empirically calibrated relationships between number and total weight of particles in given sieve fractions to obtain an approximate expression for the coefficient of variation of the proportion of a bulk GSD falling within a particular fraction, and thus a criterion for sample size for given precision. However, he neglected the stochastic variation in n and it is not obvious how to extend his approach to an analytical treatment of sampling variability in percentiles. Church *et al*. (1987) sieved replicate subsamples from a very large bulk sample of river gravel and established empirically that the proportion on a particular sieve is fairly stable once the fraction contains *c*. 100 grains; on the basis that the coarse tail is critical, they recommended that sample sizes should exceed 10^2, or preferably 10^3, times the mass of the D_{max} grain. Their approach gives a conservative guideline for reliability in the GSD generally, but it may be unnecessarily demanding if only certain percentiles are of interest.

The general lack of attention to uncertainty in GSD percentiles may stem from a feeling that sampling errors are insignificantly small for sample sizes routinely collected (typically 0·1–10kg). This may well be true for sands, since even a modest sample contains a very large number of grains (about five million in 1kg of 0·5mm sand, for example). In gravels, however, the problem is more serious for two reasons: individual grains are far larger so that a given mass contains far fewer grains, and gravels are typically far less well sorted than sands so that occasional exceptionally heavy particles may be encountered. But often the coarse percentiles which are most vulnerable to error for these reasons are the main focus of physical interest. In fluvial geomorphology, for example, D_{84} or D_{90} is used to estimate bed roughness for hydraulic calculations and mixing layer thickness in sediment routing models (e.g. Hoey and Ferguson, 1994); and inferences about whether or not bedload transport is size-selective are often based on the coarsest fractions, possibly even D_{max} (e.g. Ashworth *et al*., 1992). Sample estimates of these parameters may be strongly influenced by the presence or absence of rare large clasts. Obtaining a sample that includes large numbers of the coarsest clasts can require an overwhelming amount of work. The sample-size criteria advocated by de Vries (1970) and Church *et al*. (1987) can easily translate into samples of many tonnes. Similar practical difficulties arise with tills, debris-flow deposits, pebbly beaches, and other coarse clastic environments. In working with gravels it becomes important to have some idea of how large a sample is actually required for a given level of precision in the estimated parameters which are of most interest. In any case, we think it is good practice to report error ranges wherever possible for estimates from samples.

If a theoretical analysis of sampling variability in GSD parameters is not possible, it is necessary to estimate variability by sampling experiments. These can involve replicate sampling in the field (e.g. Mosley and Tindale, 1985; Church *et al*., 1987), replicate sampling by computer from a very large field sample treated as a population (Rice and Church, 1996), or computer sampling from an assumed population. We adopt this last approach because it is more readily extended to a wide variety of gravel GSDs covering the range of observed population characteristics, notably different degrees of spread. The obvious assumption for the population GSD is the log-normal PDF, which is a good approximation to many natural sediment GSDs. It has been suggested (e.g. Fieller *et al*., 1992) that log-hyperbolic or log-Laplace PDFs are more generally applicable, but one thorough (though as yet unpublished) study suggests that these three- and four-parameter distributions fit fluvial bed and bedload GSDs only slightly better than does the two-parameter log-normal model, the difference being mainly in the tails (M. Church, personal communication, July 1996). In this paper we present a method for computer sampling experiments on variability in percentiles of a bulk GSD, results for commonly used size percentiles as functions of sample volume when the underlying PDF is log-normal, non-dimensional generalizations of the results, and guidelines for sampling practice.

SAMPLING EXPERIMENTS

Our general approach is to obtain the sampling distributions of grain size percentiles by computer sampling experiments using replicate random samples from specified population GSDs by weight. Results are presented only for log-normal populations, but the method could be applied to other theoretical populations, or to

empirical parent GSDs as in the work of Rice and Church (1996). Experiments are done for a wide range of sample sizes, and a range of sorting coefficients as quantified by the logarithmic standard deviation σ of the population. Percentiles are for the GSD by mass, not by number as in the work of Rice and Church (1996), and are estimated in the same way as when sieving a bulk sample, i.e. by interpolation in the cumulative GSD as defined by a set of sieves rather than by knowledge of every grain.

Sampling frame

Bulk samples may be samples in space, time, or both. A sample from perhaps $1\,m^2$ of river bed or beach may be used to characterize a depositional facies extending for hundreds of square metres; a river bedload sample collected over a few minutes may be used to estimate flux over several hours. Although it is the GSD which is extrapolated from sample to population, as if it were a continuum property like, say, temperature, the fundamental assumption made is that the grains in the sample are representative of the grains in the population. Sampling experiments to investigate how well sample GSDs represent populations must therefore be based on random sampling of individual grains with a specified PDF of grain size, i.e. a GSD by number, even though the sample statistics of interest are those of the GSD by weight. To do this, the population GSD by weight must be converted to a GSD by number which is then sampled a grain at a time.

In some field situations sampling is likely to be truncated in the fine tail, coarse tail, or both. For example, when sampling a river bed beneath flowing water the finest material may be winnowed away by the current; a pit-type bedload trap may have reduced sampling efficiency for fine sediment, some of which saltates across the pit; and hand-held bedload samplers have reduced trap efficiency for grains almost as big as the sampler. Our main sampling experiments are with untruncated samples but we consider later the possible effects of truncation.

Conversion of grain size distributions

The first step in converting a population GSD by mass to one by number is to assume that all grains in the population have the same (though unspecified) density and shape. The GSD by mass is then the same as that by particle volume, particle volume is proportional to the cube of particle diameter, and the difference between true particle intermediate-axis diameter and equivalent sieve diameter does not lead to any bias. The second step is the conversion from GSD by volume to GSD by number. For convenience we define GSDs in psi units, where $\psi = \log_2(D)$; this is preferred to the more usual $\phi = -\log_2(D)$ to avoid the confusing inverse relationship between D and ϕ. Note that $D = 2^{\psi} = e^{\psi \ln(2)}$. The relationship between the PDF by volume, $f_v(\psi)$ say, and that by number, $f_n(\psi)$ say, is then:

$$f_n(\psi) = kD^{-3} f_v(\psi) = k_e^{-c\psi} f_v(\psi) \tag{2}$$

where $c = 3\ln(2)$ and k is a normalizing constant such that

$$\int_{-\infty}^{\infty} f_n(\psi)\mathrm{d}\psi = 1$$

In general, Equation 2 must be evaluated numerically and represented by a lookup table for sampling experiments, but in the important special case of a log-normal GSD by volume an analytical expression can be found for the GSD by number. If $f_v(\psi) \sim N(\mu, \sigma^2)$ then Equation 2 becomes:

$$\begin{aligned} f_n(\psi) &= \frac{k}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\psi-\mu)^2}{2\sigma^2} - c\psi \right) \\ &= \frac{k}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\psi-\mu+c\sigma^2)^2}{2\sigma^2} - c\mu + \frac{c^2\sigma^2}{2} \right) \end{aligned} \tag{3}$$

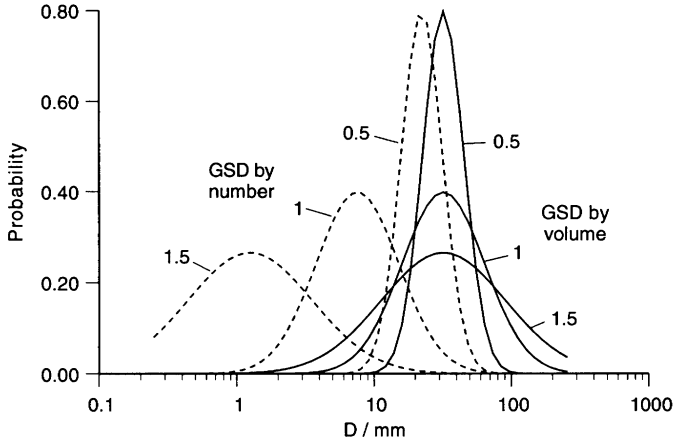

Figure 1. Log-normal grain size distributions by weight or volume (solid curves on right, centred on median diameter $D_{50}=32$ mm but with different psi standard deviations σ), compared with equivalent distributions by number (dashed curves on left). Distribution by number is shifted towards finer sizes by an amount proportional to σ^2, in accordance with Equation 3

This is a normal distribution with mean $\mu - c\sigma^2$, variance σ^2, and $k=\exp(c\mu - c^2\sigma^2/2)$. Thus the GSD by number corresponding to a log-normal GSD by volume is also log-normal, with the same standard deviation σ, but its mean is shifted by $3\sigma^2\ln(2)$ towards finer grain sizes. The magnitude of the shift increases rapidly with wider sorting of the GSD (Figure 1). It arises because grain mass is proportion to D^3; the population by number consists predominantly of small grains, but in terms of mass the rare large grains are disproportionately important so the mean by mass is much coarser than the mean by number. The less well sorted the population, the more important the few very large grains become and the greater the shift in the mean.

The range of sorting coefficients we investigated was based on our own and published bulk GSDs for gravel-bed rivers. At the Oak Creek bedload trap site in Oregon (Milhous, 1973; Parker *et al*., 1982) $\sigma=1{\cdot}07$ for the bed surface (truncated at 2 mm; $\sigma=1{\cdot}53$ including the sand tail), and 1·06 to 3·21 for bedload samples. Surface bulk samples at 11 sites along the Allt Dubhaig in Scotland (Ferguson *et al*., 1996) have σ between 0·86 and 1·53 (truncated at 0·25 mm). The North Fork Toutle River in Washington, with a patchy sand/gravel bed, has values from 1·64 to 3·05 (Seal and Paola, 1995). We chose round-number sorting coefficients of 0·5, 1 and 1·5 for the main sampling experiments (corresponding to D_{84}/D_{50} ratios of 1·4, 2·0 and 2·8) but did supplementary experiments with $\sigma=2$ and 0·25 to test generalizations made from the main experiments. Sorting coefficients >2 in the field data mentioned all relate to bimodal GSDs, generally gravel/sand mixtures; we discuss later the applicability of our results and methods to such situations.

Computational details of sampling experiments

We developed the methodology and code jointly, but coded the algorithms independently using different hardware and software (C++, QBasic) and duplicated some experiments as a check on the computations; differences in results were always small and within the limits of what would be expected by chance. A zero population mean μ (in ψ units) was used, without loss of generality since all results are relative to the mean. Sample size is defined as a total volume V of sediment (i.e. mineral volume, not bulk volume) but is always considered relative to the volume of some representative grain, generally:

$$V_{50}=\alpha D_{50}{}^3 \tag{4}$$

which is the volume of a grain of size equal to the median (D_{50}) of the population; α is a shape factor, set to $\pi/6$ in the code. Expressing sample size as V/V_{50} makes results non-dimensional and applicable to any GSD, coarse or fine, that is approximately log-normal. The main results are for $V/V_{50}=10$ to 300 000. As an indication of what this translates to in real terms, for ellipsoidal quartz-density gravel with $D_{50}=32$ mm the equivalent range of sample size is from less than 0·5 kg to nearly 15 tonnes.

The shifted normal distribution of ψ was sampled with replacement using normally distributed pseudorandom numbers generated by the Box-Muller algorithm (function GASDEV of Press *et al*., 1992) with a seed based at the start time of each simulation. In each main experiment 500 replicate random samples of a specified target size in terms of V/V_{50} were generated. Each sample was built up sequentially until its total volume exceeded the target; the grain which took the sample over the target was included or excluded depending on which choice gave a volume closer to the target. A sample was often terminated when the random number generator simulated an unusually large grain, so sample sizes fluctuated appreciably on either side of the target value, although on average they were very close to target. As each sample was built up grain volumes were accumulated separately for the range of half-psi size fractions within which percentiles of interest were expected to fall, i.e. excluding the very large numbers of grains in the fine tail. Once a sample was complete, percentiles from ψ_{50} to ψ_{99} were calculated by linear interpolation, equivalent to the standard graphical procedure of plotting percentage finer by weight against diameter (on a logarithmic scale) with points joined by straight lines. The exact diameter of the largest grain (D_{max}), total number of grains, and total sample volume were also recorded.

Run times depended on the hardware and software used, the number of replicate samples generated, and the typical number n of grains per sample. The latter increases linearly with the target sample size V/V_{50}, and much faster than linearly with the sorting coefficient σ of the population; a good empirical fit to our results is:

$$\log_{10}\left(\frac{nV_{50}}{V}\right) = 0 \cdot 95\sigma^2 \tag{5}$$

Large samples from poorly sorted populations contain upwards of 10^6 grains, which with our code and computers necessitated overnight or longer runs to obtain large numbers of replicate samples.

RESULTS

We consider two aspects of the reliability of estimated percentiles of bulk GSDs: systematic deviation between sample and population values (bias), and reproducibility of estimates in the face of random sample-to-sample scatter (precision). Sampling bias is quantified as the difference between the mean sample value of the pth percentile (ψ_p) and the equivalent population value. Precision is quantified as the between-sample standard deviation of ψ_p.

Population values of percentiles of $f_v(\psi) \sim N(\mu, \sigma^2)$ can be defined in two ways: exact, i.e. from tables of the unit normal distribution, or by interpolation in the same way as sample percentiles. Except when an exact percentile coincides with a sieve size, interpolated values of ψ_p lie on chords to the continuous curve of the exact cumulative GSD and are therefore coarser than exact percentiles where the continuous curve is concave down ($p > 50$ for the normal), but finer for those where the curve is concave up ($p < 50$). With half-psi sieve intervals the difference between exact and interpolated population percentiles is negligibly small for poorly sorted GSDs ($\sigma \geq 1 \cdot 5$), and also for percentiles even of well-sorted GSDs that happen to fall close to sieve sizes, but it exceeds 0·1psi unit for some percentiles at $\sigma = 0 \cdot 5$. We use interpolated population percentiles in order to compare like with like. However, the discrepancy between exact and interpolated percentiles must be recognized as a source of systematic error in sample GSD parameters for well-sorted populations. A special case of this discrepancy arises in the coarsest fraction of a sample, which for interpolation purposes is assumed to extend to the next sieve up; this can lead to the apparent paradox that the exact sample D_{max} is finer than the interpolated D_{99}, D_{95} or even D_{90}.

Bias in sample percentiles

Figure 2 shows how the mean difference $\Delta\psi_p$ between sample and population values of the pth percentile changes with increasing target sample size normalized as a multiple of V_{50}. Results are plotted separately for three different values of the sorting coefficient σ. Plots of median difference are very similar. Each plot shows

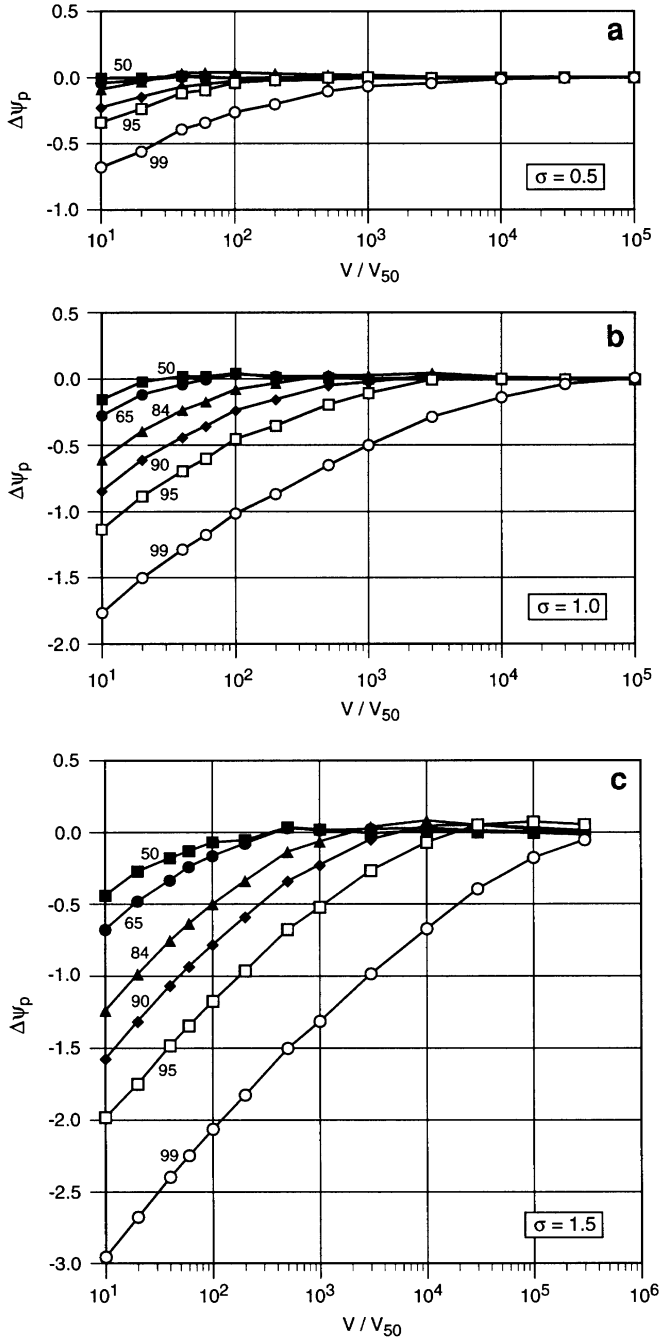

Figure 2. Relationship of bias to sample size in sample estimates of percentiles of log-normal grain size distributions with sorting coefficients σ = (a) 0·5, (b) 1, and (c) 1·5. Number by each curve denotes percentile. Bias is difference between sample-mean value in psi (negative phi) units, averaged over 500 replicate samples, and population value; negative difference indicates sample is finer than population. Sample size is a multiple of the volume V_{50} of a grain of the population median diameter

the same qualitative features: (1) it is not just D_{max} that is underestimated in small samples, but all percentiles including the median; (2) as sample size increases, the underestimation gets less and there is sometimes slight overestimation before sample percentiles converge on the population values; (3) bias continues to progressively higher sample sizes for coarser percentiles. Comparison of the three plots for different sorting coefficients σ

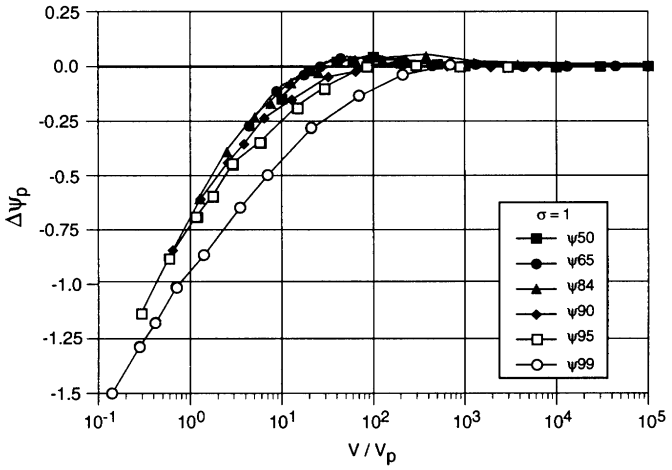

Figure 3. Data from Figure 2b replotted with sample size scaled by volume of population D_p grain size (where p is the percentile of interest). Curves for bias in different percentiles collapse close together.

shows also that (4) the amount of bias for a given percentile and sample size increases with σ, and (5) the sample size required for negligible bias in a given percentile increases with σ.

The underestimation of ψ_{50} and coarser percentiles in small samples is a consequence of the very asymmetric distribution of grain volume. Because there are far more fine particles than coarse ones, a small sample may fail to include any large grains and its D_{50} by volume or weight will be smaller than that of the population. Some small samples will contain one of the rare large particles, and thus have a much higher D_{50}, but the smaller the sample the less likely this is. This effect is more marked for coarser percentiles (since the bigger grains concerned are even rarer) and when sorting is poorer (since the coarse tail of the population GSD extends further). It may be amplified by the tendency of large grains in small samples to be rejected according to the target-size criterion discussed above, but supplementary experiments in which each sample had a randomly chosen target number of grains (rather than target volume) show identical patterns of bias. Small samples therefore tend to underestimate the median grain size and other percentiles of the bulk GSD. An individual small sample might have more or less bias than the trend in Figure 2; this variability is encompassed in the question of precision, which is discussed later.

Non-dimensional guidelines on bias

Sample size is represented in Figure 2 as V/V_{50}, which can be regarded as the number of grains there would be in the sample if they were all of size D_{50}. However, the sampling behaviour of a percentile other than the median, for example D_{84}, might be expected to depend on the equivalent number of grains of this size, not the median size, and thus on sample volume as a mulitple of (in this case) V_{84}, the volume of a D_{84}-sized grain. More generally we can define

$$V_p = \alpha D_p^3 \tag{6}$$

as the volume of a grain which falls at the pth percentile of the population GSD by weight, and investigate this as a scaling factor for sample size.

The results of Figure 2b for $\sigma = 1$ are replotted in Figure 3 as a function of V/V_p. Given that individual data points have 95 per cent confidence intervals of about $\pm 0{\cdot}05$, there is no significant difference between the curves for percentiles from ψ_{50} to ψ_{90}; ψ_{95} differs by a small amount and ψ_{99} more substantially. That is, bias in D_{84} scales on V/V_{84} in essentially the same way that bias in D_{50} scales on V/V_{50}, and so on. Unplotted percentiles between 50 and 95 would presumably follow the same trend, and a supplementary experiment in which ψ_{16} was calculated showed that it too is biased downwards at small sample sizes and follows the trend in Figure 3. For the degree of sorting in this population ($\sigma = 1$), bias in any percentile except ψ_{99} is less than $0{\cdot}15$ (a 10 per cent error in D) once the dimensionless sample size V/V_p exceeds about 20.

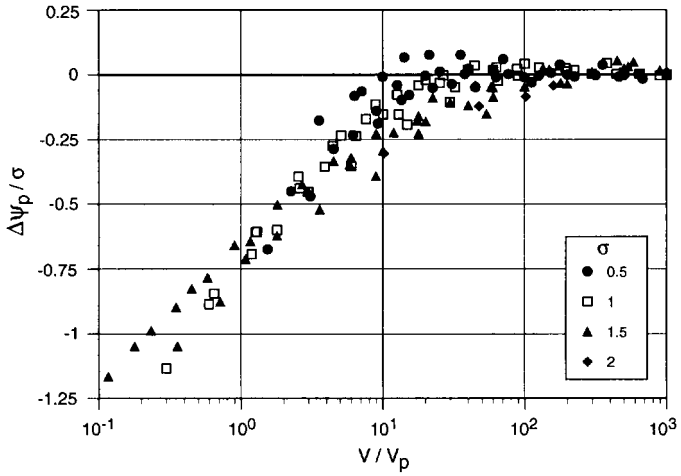

Figure 4. Approximate non-dimensional collapse of small-sample bias results (as in Figure 2) for percentiles from 50 to 95 onto a single curve for all percentiles and sorting coefficients. Sample size is scaled as in Figure 3, bias is scaled by population sorting coefficient σ

Similar plots for the populations of Figures 2a and 2c again show that all percentiles apart from ψ_{99} collapse to a single curve ($\sigma = 1{\cdot}5$) or narrow band ($\sigma = 0{\cdot}5$). However, the position of the collapse curve shifts to a small but perceptible extent with σ. With poorer sorting the bias for a given small sample size in terms of V/V_p is greater, and the sample size has to be larger for bias to become negligible and sample percentiles to converge on the true values. This is not unexpected since bias must disappear as $\sigma \rightarrow 0$ and the population becomes more and more homogeneous in grain size. It should also be noted that as σ increases, so does the ratio V_p/V_{50} for any given percentile p of interest, and thus the absolute sample size V for any given median grain size and target value of V/V_p.

Much of the difference between plots like Figure 3 for populations with different degrees of sorting is removed if bias is scaled by the sorting coefficient σ, thus making both axes of the plot non-dimensional. This is illustrated in Figure 4, which includes results from a supplementary experiment at $\sigma = 2$ as well as the main experiments. Bias reduces near-linearly with the logarithm of sample volume and disappears beyond some critical sample size V_b. This trend can be described by the relationship:

$$\Delta\psi_p / \sigma \approx 0 \cdot 5 \log_{10}(V / V_p) \tag{7}$$

fitted by eye for $V < V_b$. For $V > V_b$, $\Delta w_p \approx 0$. This collapse is not perfect, in that the data points for different sorting coefficients are slightly offset and the hinge point is really a short curve, but to a first approximation it describes the sampling behaviour of percentiles up to the 95th from any log-normal GSD with σ in the range 0·5–2 that is typical of most coarse clastic deposits. The small offset between trends for different sorting coefficients is almost eliminated by making V_b a function of σ. The relationship

$$V_b / V_p \approx 20\sigma \tag{8}$$

is a good fit for σ in the range investigated and has sensible asymptotic behaviour as $\sigma \rightarrow 0$.

Whilst scaling sample size by V_p rather than V_{50} assists generalizations about the existence and amount of bias, the difference in sample size required to avoid bias in different percentiles is more apparent if V/V_{50} is used. Equations 7 and 8 can be expressed in terms of V/V_{50} by noting that for a log-normal population of similar-shaped grains:

$$\frac{V_p}{V_{50}} = \left(\frac{D_p}{D_{50}}\right)^3 = 2^{3\sigma z_p} \tag{9a}$$

where z_p is the pth percentile point of the unit normal distribution. This is equivalent to:

$$\log_{10}\left(\frac{V_p}{V_{50}}\right) = 3\sigma z_p \log_{10}(2) = 0 \cdot 9\sigma z_p \tag{9b}$$

The critical sample size for negligible bias, V_b, is then defined by:

$$\log_{10}(V_b / V_{50}) \approx 1 \cdot 3 + \log_{10}(\sigma) + 0 \cdot 9\sigma z_p \tag{10}$$

and the extent of bias at smaller sample sizes becomes:

$$\Delta\psi / \sigma \approx 0 \cdot 5 \log_{10}(V / V_{50}) - 0 \cdot 65 - 0 \cdot 5 \log_{10}(\sigma) - 0 \cdot 45\sigma z_p \tag{11}$$

These relationships provide a good fit to the data of Figure 4, i.e. for sorting coefficients from 0·5 to 2. The overall root-mean-square error in predicting $\Delta\psi_p$ is 0·10 (a 7 per cent error in D) and differs little between experiments. A supplementary sampling experiment with a very well-sorted population ($\sigma = 0 \cdot 25$) showed bias in very small samples, with the existence of bias predicted fairly well by Equation 8, though Equation 11 tended to overestimate its magnitude. Our results therefore apply to a wide range of well to poorly sorted conditions, so long as the GSD is unimodal and approximately log-normal. They provide a basis for assessing the conditions under which, and if so the extent to which, sample GSD percentiles are likely to underestimate the equivalent population values.

Precision of percentiles

Even when sample sizes are sufficient to avoid significant bias, an individual sample may deviate substantially from the population GSD. This lack of precision can be quantified using the standard deviation s_p of ψ_p in replicate samples. If there is no bias, s_p is the same as the root-mean-square (r.m.s.) deviation of sample values from the true value, but if bias is present the r.m.s. deviation is larger than s_p because it combines the random and systematic variance. In situations where sampling is random and the central limit theorem applies, the standard deviation of many sample statistics decreases as the square root of sample size; we shall refer to this as $1/\sqrt{n}$ behaviour. As already noted, it is theoretically expected, and empirically observed, when estimating GSDs by pebble count methods. In bulk GSDs n is not known but a dimensionless sample size such as V/V_{50} or V/V_p might be expected to play a similar role and give what may be termed $1/\sqrt{V}$ behaviour.

Figure 5 shows how s_p varies with sample size expressed as V/V_{50} in experiments with three different values of the population sorting coefficient σ. The plots show substantial sample-to-sample variability in GSD percentiles and, as expected, precision is worse in more poorly sorted populations. Over much of the range of experimental conditions $s_p > 0 \cdot 3$, implying 95 per cent confidence intervals wider than $\pm 0 \cdot 6$ psi (a + 50 per cent, –33 per cent factor of uncertainty in D_p, rising to a factor for two for coarse percentiles of poorly sorted GSDs). Interestingly, there is widespread deviation from $1/\sqrt{V}$ behaviour. Only the curve for ψ_{50} in the best-sorted case shows s_p declining as the inverse square root of V/V_{50} throughout the plotted range of sample size. In other cases $1/\sqrt{V}$ behaviour does not set in until quite large sample sizes, and at smaller sample sizes s_p is roughly constant or even increases with sample size (e.g. ψ_{50} at $\sigma = 1$ and 1·5). In one case (ψ_{99} with $\sigma = 1 \cdot 5$) there is no improvement in precision throughout the 10^4 range of sample sizes considered. The onset of $1/\sqrt{V}$ behaviour is delayed to larger sample sizes when the population is more poorly sorted.

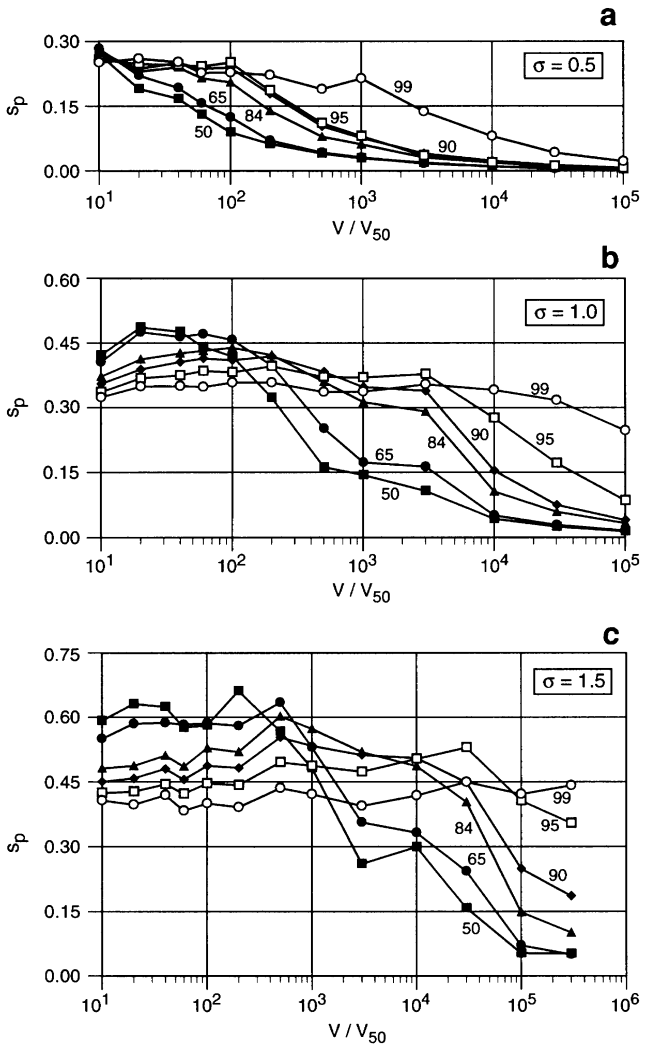

Figure 5. Changing precision of sample estimates of percentiles as sample volume increases. Number by each curve denotes percentile. Ordinate is standard deviation in psi units of 500 replicate sample estimates, plotted as in Figure 2 against sample size as a multiple of volume of median-sized grain, separately for different population sorting coefficients σ = (a) 0·5, (b) 1, and (c) 1·5. Note that (c) covers one more log cycle of sample size than the others

Non-dimensional guidelines on precision

As in the case of bias, it seems reasonable to expect the precision of sample percentiles to depend on sample size relative to the grain volume V_p corresponding to the percentile concerned rather than the volume V_{50} of a median-sized grain. Standard sampling theory also suggests s_p will scale directly with σ. A combined plot of the data of Figure 5 together with those for σ = 2, with V normalized by V_p and s_p normalized by σ, is shown in Figure 6.

For any given population this scaling appears to collapse differences between percentiles onto a similar two-part trend: near-horizontal at some constant standard deviation s_0 for sample sizes up to a critical value V_c, then approximately linear downwards with slope –0·5, indicating the $1/\sqrt{V}$ behaviour anticipated for all sample sizes. The second phase can be represented by:

$$\frac{s_p}{s_0} = \left(\frac{V}{V_c}\right)^{-0\cdot5} \tag{12}$$

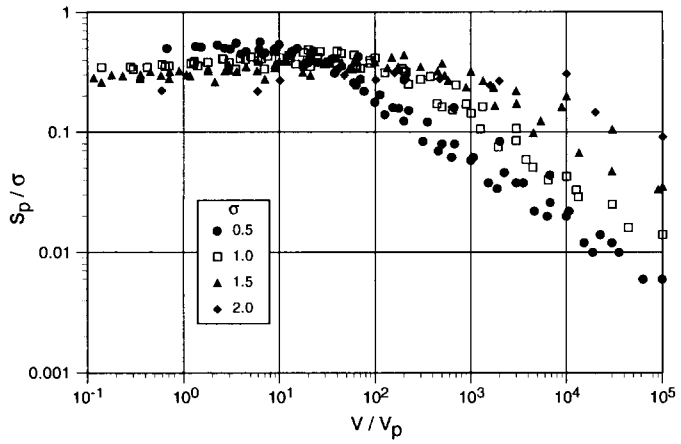

Fig. 6. Precision results as in Figure 5 replotted in non-dimensional format of Figure 4, with precision scaled by sorting coefficient σ and sample volume scaled by volume of a grain of the corresponding population percentile size. With logarithmic scales on both axes, $1/\sqrt{n}$ sampling error is represented by a linear trend with slope –0·5

However, when different populations are compared it is evident that the critical sample size V_c and constant-precision value s_0 both increase with σ. The approximations

$$s_0 \approx 0 \cdot 4\sigma^{0 \cdot 6} \tag{13}$$

and

$$\frac{V_c}{V_p} \approx 160\sigma^3 \tag{14}$$

give a good visual fit and have asymptotically sensible behaviour as $\sigma \to 0$. The four two-part trend lines given by Equations 12–14 with $\sigma = 0 \cdot 5$, 1, 1·5, and 2 are omitted from Figure 6 for clarity but the deviations from them are small, with a r.m.s. error (for s_p/σ) of 0·04 in each experiment. Precision results from the supplementary experiment with $\sigma = 0 \cdot 25$ are predicted quite well by Equations 12–14, with a r.m.s. error (for s_p) of 0·03 psi units. For all realistic values of σ, $V_c > V_b$; that is, the precision of sample percentiles does not start to improve until a critical sample size which is greater than that for negligible bias.

These generalized relationships for precision as a function of V/V_p can be expressed in terms of V_{50} instead by using Equation 9. The critical sample size for the onset of improved precision and $1/\sqrt{V}$ behaviour then becomes:

$$\log_{10}\left(\frac{V_c}{V_{50}}\right) \approx 2 \cdot 2 + 3\log_{10}(\sigma) + 0 \cdot 9\sigma z_p \tag{15}$$

When $V > V_c$ the expectable precision is, from Equations 12 and 15:

$$\log_{10}\left(\frac{s_p}{s_0}\right) = 1 \cdot 1 + 1 \cdot 5\log_{10}(\sigma) + 0 \cdot 45\sigma z_p - 0 \cdot 5\log_{10}\left(\frac{V}{V_{50}}\right) \tag{16}$$

or, using Equation 13 as an empirical approximation for s_0:

$$\log_{10}(s_p) = 0 \cdot 7 + 2 \cdot 1\log_{10}(\sigma) + 0 \cdot 45\sigma z_p - 0 \cdot 5\log_{10}\left(\frac{V}{V_{50}}\right) \tag{17}$$

These results quantify the approximate precision expected in estimating any percentile between D_{50} and D_{99} of an approximately log-normal GSD with sorting from very good ($\sigma = 0{\cdot}5$) to poor ($\sigma = 2$). In particular, Equation 14 or 15 indicates the sample size required for precision to start improving in the way generally expected in random sampling. Depending on whether a particular sample is below or above this critical size, Equation 13 or 17 then indicates the likely precision in ψ_p. The value of s_p can be doubled to act as an approximate 95 per cent confidence interval for an estimate. As discussed next, Equation 17 can also be used to assess the sample size required in order to estimate a particular percentile to some required precision.

IMPLICATIONS FOR SAMPLE SIZE

The few published guidelines on sample sizes to characterize gravel GSDs (de Vries, 1970; ISO, 1977; Church *et al.*, 1987; Gale and Hoare, 1992) involve some multiple of the mass or volume of the D_{84} or $D_{\max}$ grain, irrespective of the purpose of sampling. Our results show that different criteria are required depending on which GSD percentile is of interest, and whether the only consideration is to avoid bias or also to achieve a specified precision better than that which holds at the minimum sample size for negligible bias.

The finding that all percentiles, including the median, are likely to be underestimated in small samples points to a minimal criterion that samples should be big enough to avoid significant bias in the percentile or percentiles of interest. Equations 8 or 10 indicate the sample size V_b required to achieve this. It increases with the sorting coefficient σ of the GSD and also for percentiles further into the coarse tail of the distribution. These effects are multiplicative in the last term of the equation, so the sample size for negligible bias in D_{50} must be increased by a factor of almost 10 to avoid bias in D_{84} when $\sigma = 1$, but by a factor of 30 when $\sigma = 1{\cdot}5$.

Once sample sizes exceed V_b bias is negligible, but estimated percentiles remain imprecise and precision stays constant or gets slightly worse for sample sizes up to the higher sample size V_c. Only for sample sizes beyond V_c does precision improve as $1/\sqrt{V}$ in the way generally expected. Thus there is no benefit in increasing sample size by a small amount above V_b; more precise estimates require a sample size some way above V_c. The sample size required to achieve a specified level of precision for a particular percentile p and population sorting coefficient σ can be obtained by inverting Equation 17:

$$\log_{10}\left(\frac{V}{V_{50}}\right) = 1{\cdot}4 + 4{\cdot}2\log_{10}(\sigma) + 0{\cdot}9\sigma z_p - 2\log_{10}(s_p) \tag{18}$$

This allows sample sizes to be planned using a second and more stringent criterion: that the sample estimate of a grain-size percentile should not only be unbiased but should also have a prescribed, small, standard error. For illustrative purposes we use $s_p = 0{\cdot}15$ (corresponding to a 10 per cent standard error in D_p) in Equation 18 to define a sample size V_g for 'good' precision:

$$\log_{10}\left(\frac{V_g}{V_{50}}\right) = 3{\cdot}0 + 4{\cdot}2\log_{10}(\sigma) + 0{\cdot}9\sigma z_p \tag{19}$$

As with the no-bias criterion, the required sample size for good precision increases with three factors: average grain size (as represented by V_{50}), scatter of grain sizes (σ), and percentile of interest (through z_p).

The sample sizes required to avoid bias and to achieve good precision are plotted in Figure 7 as a function of the sorting coefficient σ, with separate lines for the widely used 50th, 84th and 95th percentiles. The lines for coarser percentiles are steeper because of the multiplicative effect of p and σ. The lines for good precision plot higher than those for no bias because V_g (Equation 19) exceeds V_b (Equation 10), more so for more poorly sorted GSDs.

Figure 7 also shows the two best-known general-purpose guidelines for sample size. De Vries (1970), whose work is the basis for ISO (1977), used a binomial model to deduce the sample size necessary to estimate the GSD proportion on one sieve to a relative precision r. His formula for 'normal' accuracy of $r = 0{\cdot}03$ in fractions around D_{84} and comprising c. 10 per cent of the GSD amounts in our notation to:

$$\log_{10}\left(\frac{V}{V_{84}}\right) \approx 4 \tag{20}$$

Since $z_{84}=1{\cdot}0$, this is equivalent using Equation 9b to:

$$\log_{10}\left(\frac{V}{V_{50}}\right) \approx 4 + 0{\cdot}9\sigma \tag{21}$$

The previously mentioned recommendation of Church *et al*. (1987) that the D_{max} grain should constitute less than 1 per cent of the sample, or ideally 0·1 per cent, is equivalent to $V/V_{max}=100$ or 1000. If D_{max} is approximated by the population D_{99} the 1 per cent criterion gives:

$$\log_{10}\left(\frac{V}{V_{50}}\right) \approx 2 + 2{\cdot}1\sigma \tag{22}$$

with the constant 2 becoming 3 for a 0·1 per cent criterion. As can be seen from Figure 7, the guidelines of de Vries and Church differ from each other and from those we propose. Both suggest sample sizes far in excess of

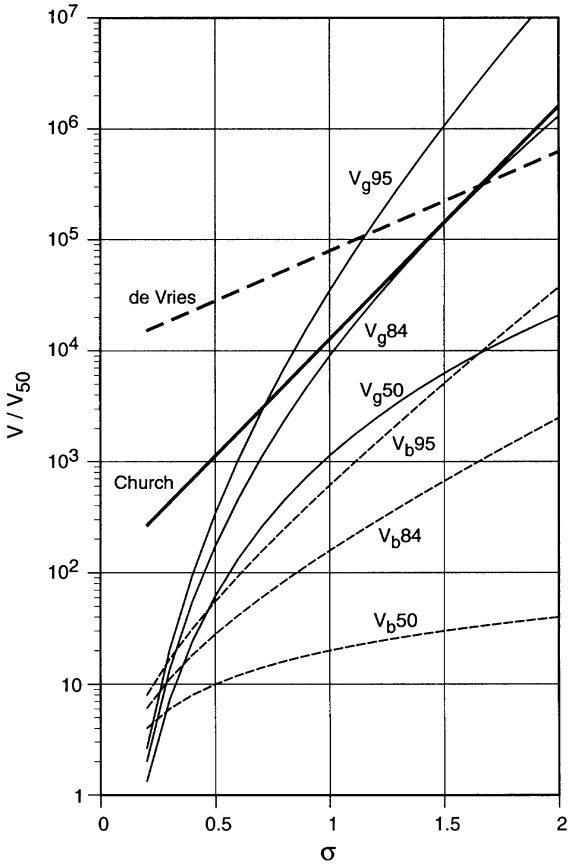

Figure 7. Necessary sample sizes for negligible bias (V_b, Equation 10) and precision of 0·15 psi units or 10 per cent of diameter (V_g, Equation 19). Sample size is represented non-dimensionally as a multiple of the volume V_{50} of a median-sized grain. Included for comparison are more general sample-size guidelines from de Vries (1970; 'normal accuracy') and Church *et al*. (1987; D_{max} <1 per cent of sample). See text for details and discussion

those needed to avoid the small-sample bias identified in our simulation experiments. With well-sorted GSDs the previous guidelines, particularly that of de Vries, also suggest sample sizes much higher than we find necessary to achieve good precision. For σ in the range 0·5 to 1 our criterion of a 10 per cent standard error in D_p is achieved with samples one to two orders of magnitude smaller than de Vries' criterion would suggest (the higher value being when the coarse tail of the GSD is not of interest), and up to an order of magnitude smaller than Church's 1 per cent criterion would suggest. However, gravel GSDs are rarely so well sorted. As σ increases the guidelines of de Vries and Church converge and eventually cross, and some members of our family of guidelines converge or cross one or both of the published guidelines. With poorly sorted GSDs our results suggest that the required sample size varies greatly depending on which percentile is of interest: good precision in the sample median is achieved with sample sizes at least an order of magnitude smaller than suggested by the published guidelines, but good precision in D_{95} requires samples as big or bigger than previously suggested. It is important to recognize that de Vries (1970) and Church *et al*. (1987) were proposing guidelines for a sample size adequate to characterize the entire GSD, including the coarse tail, so it is not surprising that our results suggest that smaller samples are adequate for the simpler task of estimating a percentile that is not in the coarse tail.

Figure 7 is non-dimensional, with required sample volume V scaled by the volume V_{50} of a median-sized grain. It can be converted to sample mass using Equation 4 for V_{50} (which involves a shape factor) and an assumption or measurement of sediment density. For illustrative purposes we take the case of ellipsoidal quartz-density grains with $D_{50} = 32$ mm. A median-sized grain then weights 0·045 kg, the vertical axis of Figure 7 extends from this value to 450 tonnes, and a 1000-kg sample corresponds to $V/V_{50} \approx 20\,000$. For a moderately well-sorted GSD ($\sigma = 1$), bias in D_{50}, D_{84} and D_{95} becomes negligible in samples exceeding about 1, 7 and 30 kg, respectively. Good precision in these percentiles requires samples of order 50, 400 and 1600 kg. For comparison, the 'normal accuracy' guideline of de Vries (1970) and 1 per cent criterion of Church *et al*. (1987) give target sizes of about 3600 kg and 600 kg. Typical fieldwork practice when sampling coarse clastic deposits would meet the no-bias requirements easily, and probably also the good-precision criterion for D_{50} and D_{84}, but often not that for D_{95}, and sampling of sediment fluxes may often fail to meet the criteria. Fluvial geomorphology can be taken as a specific example of an application area. Bedload sampling over the full width of a channel using a pit or vortex-tube extractor would generally meet the no-bias criterion and achieve good precision for D_{50}, but often not for D_{84}. Bedload sampling at a point using a portable sampler might meet the no-bias criterion for D_{50}, but might well underestimate tail percentiles and would not give good precision (as we have defined it) for any percentile. For a more poorly sorted GSD ($\sigma = 1{\cdot}5$) with $D_{50} = 32$ mm, the required sample sizes to avoid bias increase to about 2, 30 and >200 kg for D_{50}, D_{84} and D_{95}, respectively, and those for good precision in these percentiles increase even more to about 300 kg, >6 tonnes, and nearly 50 tonnes, respectively. Obtaining and sieving such large samples of bed material requires a major effort and considerable time; doing it for bedload will seldom be possible.

DISCUSSION

Real-life sediment analysis can involve many complications not addressed in this paper. We consider only bulk GSDs, not the conversion of pebble-count or photographic GSDs to their bulk equivalents (e.g. Fraccarollo and Marion, 1995). We do not consider the trap efficiency of devices for obtaining bulk samples, although we recognize that this can be a serious problem, particularly when sampling sediment flux rather than deposits, and therefore discuss below the applicability of our method and results to truncated samples. We assume that samples are drawn from homogeneous populations and therefore sidestep questions of defining and sampling the coarse surface layer of a deposit (e.g. Fripp and Diplas, 1993), how best to characterize the overall GSD of a spatially variable deposit (e.g. Mosley and Tindale, 1985; Church *et al*., 1987; Wolcott and Church, 1991), or how fluvial bedload transport fluctuates over time (e.g. Gomez, 1991). The issue then reduces to how well the GSD of a bulk sample represents the population from which it is drawn, and what lessons can be learnt about desirable sample size.

The three key results of our sampling experiments are that (1) grain size percentiles are underestimated in small samples, (2) only after a certain, fairly large, sample size is reached does sampling variability in a given

Table I. Equation references for calculating likely bias and precision (in psi units: $\log_2$ of diameter) of any percentile p of a bulk-sieved grain size distribution

Bias of estimated percentile	
Equation 11 if $V<V_b$ (from Equation 10)	
zero if $V>V_b$	
Precision of estimated percentile	
Equation 13 if $V<V_c$ (from Equation 15)	
Equation 17 if $V>V_c$	
Minimum sample size to estimate percentile	
without bias: Equation 10	
to specified precision: Equation 18	

Calculations require an estimate of the volume V_{50} of a median-sized grain (e.g. from Equation 4 and the sample D_{50}), an estimate of the psi standard deviation σ of the grain size distribution (e.g. from estimated 16th and 84th percentiles), and the normal distribution z value corresponding to the percentile p of interest

percentile decrease, and (3) bias and precision depend on the sorting coefficient σ. The less well sorted the GSD, the greater the small-sample bias and variability, and the greater the sample sizes needed to avoid bias and improve precision. Taken together, these findings mean that no simple guideline about desirable sample size is possible: it depends on what GSD percentile is of interest, what the distribution is like, and what level of accuracy is desired. Nevertheless, by non-dimensionalizing sample size and properties we have been able to establish general trends (Figure 4 and 6) and predictive relationships (Equations 7–11, 12–16) for bias and precision and thus to propose a family of guidelines for sample size (Figure 7, Equations 10 and 19). The key equations for different purposes are listed in Table I.

The systematic bias in small-sample estimates of bulk-GSD percentiles, and the failure of precision to improve as $1/\sqrt{V}$ until fairly large sample sizes are exceeded, are previously unreported features and differ from what has been found in studies of pebble-count GSD statistics (e.g. Rice and Church, 1996). The difference is not a consequence of different sampling constraints (fixed n but stochastically varying V in pebble counting, near-fixed V but varying n in our experiments), because our supplementary experiments with different fixed n show the same pattern of bias. The most convincing explanation for bias remains the high skewness of the PDF of particle mass and thus the likelihood that small samples will systematically fail to include a representative number of the relatively rare large clasts that define the coarse tail of the bulk GSD. The near-constant precision at small to medium sample sizes is also linked to this. In small samples the sample-mean grain size is lower, and variance might then be expected also to be lower than it would otherwise be. To put it a different way, because small samples systematically under-represent the coarse tail of the population, the effective sorting coefficient is lower (D_{84}/D_{50} averages only half its population value in some of our small-sample experiments), and precision for a given sample size is therefore greater. However, this link with bias cannot be the whole explanation because precision stays near-constant over a two- to 20-fold range of sample sizes between V_b and V_c

The dependence of bias, precision and guideline sample size on the sorting coefficient is not surprising. We have argued that sample bias and imprecision stem from large sample-to-sample variability in representation of the coarse tail of the population GSD. The volume or mass of a coarse-tail grain, relative to that of a median-sized grain, increases as σ^3. As σ increases, the GSD by number is shifted much further relative to the GSD by mass (Figure 1) and the coarse grains that determine percentiles from the median upwards become even rarer, with more scope for sampling variability. The scaling of bias and precision by σ in the collapse plots of Figures 4 and 6 deals with this to some extent, but not perfectly because the relationships between σ and the number of grains in the total sample and in individual sieve fractions are highly non-linear; this is reflected in the form of the empirical collapse of Equation 13. Previous studies have not quantified the effect of sorting but have implicitly or explicitly recognized its relevance. The sample-size guidelines proposed by de Vries (1970) and

Church *et al*. (1987) used multiples of the mass of the D_{84} or D_{max} grain, respectively, and for a given D_{50} these masses vary as σ^3. Gale and Hoare (1992, 1994), in an investigation of how the results underlying the guideline of Church *et al*. (1987) varied in sediments of non-fluvial origin, noted that beach gravels behaved much like the fluvial sediments analysed by Church *et al*. but that desirable minimum sample sizes would be greater for till and other poorly sorted deposits.

A limitation, albeit inevitable, of our results is that they are generalized by scaling sample volume V as a multiple of an unknown population parameter, the volume V_p of a grain at the percentile p of interest (or equivalently, the volume V_{50} of a median-sized grain in the population together with the population sorting coefficient σ). These parameters are not known *a priori*. When assessing the precision of sample GSD percentiles, V_{50} and σ must first be estimated from the sample. When deciding required sample size in advance of a field investigation it is necessary to estimate V_{50} from a pilot study, then either estimate σ or assume a value (or range of values) by analogy with sites with known GSDs. A similar problem applies to existing guidelines, which assume a knowledge of D_{84} (de Vries) or D_{max} (Church *et al*). Further research is needed on how to estimate σ, since our results suggest that ψ_{84}/ψ_{50} will underestimate it in a small sample; a possible approach is to assume that ψ_{max} represents the xth percentile of a normal population, where x will be in the nineties and increase with sample size.

Another limitation is that our results are for untruncated samples from a log-normal population which has infinite tails. As previously discussed, many sediments deviate from a log-normal GSD, particularly in the tails which may be truncated because of thresholds in transport domains (e.g. between no motion, traction, and suspension). Even if the population is log-normal the techniques used to obtain bulk samples may impose partial or complete truncation in one or both tails. The extent to which the sampling behaviour of GSD statistics is unchanged when either the population, or sampling, is truncated deserves detailed examination. Preliminary experiments with truncated samples from a log-normal population show that bias remains, and is bigger in smaller samples, but its magnitude is slightly reduced for any given sample size. Investigation of sampling variability when populations depart drastically from log-normality could be done by computer sampling experiments of the type reported here, but using the general by-mass to by-number conversion of Equation 2 rather than the specific version (Equation 3) for the log-normal distribution. For strongly bimodal distributions, our results could be applied approximately to higher percentiles by treating the gravel mode as a separate log-normal population, assuming that sampling errors in the gravel/sand ratio and sand-mode GSD are small compared to those in the gravel-mode GSD, and adjusting percentiles and sample sizes according to the gravel/sand ratio. Thus for a sample of a sediment comprising S per cent sand and G per cent gravel, with the gravel mode log-normal, the overall D_{84} is the $100(84–S)/G$ percentile of the gravel and its sampling error would be estimated using σ for the gravel mode only and V/V_{50} based on G per cent of the total sample volume V_{50} from the gravel median, which is $D_{S+G/2}$ of the overall GSD.

Our approximate quantification of expected bias and precision in percentiles of bulk GSDs allows a new set of guidelines to be presented for the sample size necessary for reasonably accurate estimation of grain size percentiles, supplementing the more general guidelines of de Vries (1970) and Church *et al*. (1987). The recognition that small samples give biased estimates means that two guidelines are needed, one for the sample size to avoid significant bias and another for the (much larger) sample size to achieve some required precision in the estimated percentile. Moreover, both bias and precision depend on the sorting coefficient σ, and on which percentile is being estimated, so that families of guidelines become necessary rather than a single rule of thumb. We have attempted in Figure 7 to present a widely useful set of curves for the two different criteria (negligible bias, good precision), three widely used percentiles, and the range of sorting coefficients likely to be encountered. Because sample size is made non-dimensional the chart can be applied irrespective of average grain size.

The ease of achieving the target sample sizes indicated by Figure 7 depends on the average grain size, the sorting coefficient, and the type of sampling involved. If the population to be characterized is coarse and poorly sorted, the sample size required for good precision in coarse-tail percentiles can be very large indeed, as indicated at the end of our Results section for the case of $D_{50} = 32$ mm and $\sigma = 1{\cdot}5$. The other situation in which Figure 7 has alarming implications is the estimation of bedload grain size statistics from small samples obtained in hand-held sampling devices. If the bedload is gravel, sample V/V_{50} is unlikely to exceed 10^2–10^3 and grain

size percentiles will not only have poor precision but may also be biased downwards; this could be misleading in studies of the selectivity of bedload transport, especially in attempts to establish transport similarity or equal mobility. However, the results presented here still permit an assessment of the likelihood of such bias and its expected magnitude.

ACKNOWLEDGEMENTS

We thank Mike Church for information on the fit of the log-normal and other distributions to field data. He and Trevor Hoey also made helpful review comments on context and presentation. Our interest in this problem stemmed from field measurements of bedload transport and downstream fining, with support from the Royal Society, UK Natural Environment Research Council (GR3/7407), US Geological Survey, and US National Science Foundation (EAR 9004756 and 9205817).

REFERENCES

Ashworth, P. J., Ashmore, P. E., Ferguson, R. I., Paola, C., Powell, D. M. and Prestegaard, K. L. 1992. 'Measurements in a braided river chute and lobe: II Selective entrainment, transport and deposition', *Water Resources Research*, **28**, 1877–1886.

Church, M. A., McLean, D. G. and Wolcott, J. F. 1987. 'River bed gravels: sampling and analysis', in Thorne, C. R., Bathurst, J. C. and Hey, R. D. (Eds), *Sediment Transport in Gravel-bed Rivers*, Wiley, Chichester, 43–88.

Ferguson, R., Hoey, T., Wathen, S. and Werritty, A. 1996. 'Field evidence for rapid downstream fining of river gravels through selective transport', *Geology*, **24**, 179–182.

Fieller, N. R. J., Flenley, E. C. and Olbrecht, W. 1992. 'The statistics of particle size data', *Applied Statistics*, **41**, 127–146.

Fraccarollo, L. and Marion, A. 1995. 'Statistical approach to bed-material surface sampling', *Journal of Hydraulic Engineering*, **121**, 540–545.

Fripp, J. H. and Diplas, P. 1993. 'Surface sampling in gravel streams', *Journal of Hydraulic Engineering*, **119**, 473–490.

Gale, S. J. and Hoare, P. G. 1992. 'Bulk sampling coarse clastic sediments for particle-size analysis', *Earth Surface Processes and Landforms*, **17**, 729–734.

Gale, S. J. and Hoare, P. G. 1994. 'Reply: bulk sampling of coarse clastic sediments for particle-size analysis', *Earth Surface Processes and Landforms*, **19**, 263–268.

Gomez, B. 1991. 'Bedload transport', *Earth-Science Reviews*, **31**, 89–132.

Hoey, T. B. and Ferguson, R. I. 1994. 'Numerical simulation of downstream fining by selective transport in gravel-bed rivers: model development and illustration', *Water Resources Research*, **30**, 2251–2260.

International Standards Organisation. 1977. 'Liquid flow measurement in open channels – bed material sampling', *ISO-4364-1977 (E)*, 13 pp.

Milhous, R. T. 1973. *Sediment transport in a gravel-bottomed stream*, PhD thesis, Oregon State University.

Mosley, M. P. and Tindale, D. S. 1985. 'Sediment variability and bed material sampling in gravel-bed rivers', *Earth Surface Processes and Landforms*, **10**, 465–482.

Parker, G., Klingeman, P. C. and McLean, D. G. 1982. 'Bedload and size distribution in paved gravel-bed streams', *Journal of the Hydraulics Division ASCE*, **108**, 544–571.

Press, W. H., Flannery, B. P., Tuekolsky, S. A. and Vetterling, W. T. 1992. *Numerical Recipes in C*, 2nd edn, CUP, Cambridge, 964 pp.

Rice, S. and Church, M. 1996. 'Sampling surficial fluvial gravels: the precision of size distribution percentile estimates', *Journal of Sedimentary Research* A, **66**, 654–665.

Seal, R. and Paola, C. 1995. 'Observations of downstream fining on the North Fork Toutle River near Mount St Helens, Washington', *Water Resources Research*, **31**, 1409–1419.

de Vries, M. 1970. 'On the accuracy of bed-material sampling', *Journal of Hydraulics Research*, **8**, 523–533.

Wilcock, P. R. 1992. 'Flow competence: a criticism of a classic concept', *Earth Surface Processes and Landforms*, **17**, 289–298.

Wolcott, J. and Church, M. 1991. 'Strategies for sampling spatially heterogeneous phenomena: the example of river gravels', *Journal of Sedimentary Petrology*, **61**, 534–543.

Yule, G. U. and Kendall, M. G. 1953. *An Introduction to the Theory of Statistics*, 14th edn, Griffin, London, 701 pp.

